CHAPTER 4

Stationary Stochastic Processes

4.1. Weak Definition of a Stochastic Process

This section is devoted to further topics in the theory of stochastic processes and of their applications. We start with a different, weaker, definition of a stochastic process, useful in the study of stationary processes.

Consider a collection of random variables $u(t,\omega) \in \mathbb{C}$ parametrized by t.

DEFINITION. We say that $u(t,\omega)$ is a weakly defined stochastic process if for every finite set of points $t_1,\dots,t_n$ the joint distribution of $u(t_1,\omega)$, ..., $u(t_n,\omega)$ is known

$$F_{t_1,\dots,t_n}(y_1,\dots,y_n) = P(u(t_1) \le y_1,\dots,u(t_n) \le y_n)$$

The family of functions $F_{t_1,\dots,t_n}(y_1,\dots,y_n)$ must satisfy some natural requirements:

(1) $F \ge 0$.
(2) $F(\infty,\dots,\infty) = 1$ and $F(-\infty,\dots,-\infty) = 0$.
(3) $F_{t_1,\dots,t_n}(y_1,\dots,y_m,\infty,\dots,\infty) = F_{t_1,\dots,t_m}(y_1,\dots,y_m)$.
(4) If $(i_1,\dots,i_n)$ is a permutation of $(1,\dots,n)$, then

$$F_{t_{i_1},\dots,t_{i_n}}(y_{i_1},\dots,y_{i_n}) = F_{t_1,\dots,t_n}(y_1,\dots,y_n).$$

A moment of $u(t,\omega)$ of order q is an object of the form

$$M_{i_1,\dots,i_n} = E[u^{i_1}(t_1)\dots u^{i_n}(t_n)], \quad \sum_{j=1}^{n} i_j = q.$$

If a stochastic process has finite moments of order q it is process of order q. The moment

$$E[u(t,\omega)] = m(t)$$

is the mean of u at t. The function

$$E[(u(t_2,\omega) - m(t_2))\overline{(u(t_1,\omega) - m(t_1))}] = R(t_1,t_2)$$

is the covariance of u. Let us list the properties of the covariance of u:

(1) $R(t_1,t_2) = \overline{R(t_2,t_1)}$.

(2) $R(t_1, t_1) \geq 0$.
(3) $|R(t_1, t_2)| \leq \sqrt{R(t_1, t_1)R(t_2, t_2)}$.
(4) For all $t_1, \ldots, t_n$ and all $z_1, \ldots, z_n \in \mathbb{C}$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} R(t_i, t_j) z_i \overline{z_j} \geq 0.$$

The first three properties are easy to establish; the fourth is proved as follows: For any choice of complex numbers z_j, the sum

$$\sum_{i=1}^{n}\sum_{j=1}^{n} R(t_i, t_j) z_i \overline{z_j}$$

is by definition equal to

$$= E\left[\left|\sum_{j=1}^{n} \left(u(t_j) - m(t_j)\right) z_j\right|^2\right] \geq 0,$$

i.e., to the expected value of a non-negative quantity.

Definition. A weakly defined stochastic process is stationary in the strict sense if for every $t_1, \ldots, t_n$ and for any $T \in \mathbb{R}$

$$F_{t_1,\ldots,t_n}(y_1, \ldots, y_n) = F_{t_1+T,\ldots,t_n+T}(y_1, \ldots, y_n).$$

For a stochastic process which is stationary in this sense all moments are constant in time, and in particular $m(t) = m$ and $R(t_1, t_2) = R(t_1+T, t_2+T)$ for all T. Choose $T = -t_1$; then $R(t_1, t_2) = R(0, t_2-t_1)$ and it becomes reasonable to define

$$R(t_2 - t_1) = R(0, t_2 - t_1),$$

where the function R on the left side, which has only one argument, is also called R with the hope that there is no ambiguity.

The above properties become, for the new function R:

(1) $R(t) = \overline{R(-t)}$.
(2) $R(0) \geq 0$.
(3) $|R(t)| \leq R(0)$.
(4) For all $t_1, \ldots, t_n$ and all $z_1, \ldots, z_n \in \mathbb{C}$

$$\sum_{i}^{n}\sum_{j}^{n} R(t_j - t_i) z_i \overline{z_j} \geq 0. \tag{4.1}$$

Definition. A stochastic process is stationary in the wide sense if it has a constant mean and its covariance depends only on the difference between the arguments, i.e.,

(1) $m(t) = m$.

(2) $R(t_1, t_2) = R(t_2 - t_1)$.

Note that if a stochastic process is stationary in the weak sense and Gaussian then it is stationary in the strict sense (because a Gaussian process is fully determined by its mean and covariances). Brownian motion is *not* stationary. White noise is stationary (but ill-defined without appeal to distributions).

Let us find examples of stationary stochastic processes. Pick $\xi \in \mathbb{C}$ to be a random variable, $f(t)$ a real, non-random function of time, and consider $u(t, \omega) = \xi f(t)$. Assume for simplicity that $f(t)$ is differentiable. We determine when a process of this type is stationary in the wide sense. Its mean is

$$m(t) = E[\xi f(t)] = f(t)E[\xi]$$

which is constant if and only if $f(t)$ is constant or $E[\xi] = 0$. The covariance

$$R(t_1, t_2) = E[\xi f(t_2)\overline{\xi f(t_1)}] = E[\xi\overline{\xi}]f(t_2)\overline{f(t_1)}$$

must depend only on the difference $t_2 - t_1$. Consider the special case $t_1 = t_2 = t$. In this case the covariance $E[\xi\overline{\xi}]f(t)\overline{f(t)}$ must be $R(0)$ hence $f(t)\overline{f(t)}$ must be constant. Therefore $f(t)$ is of the form

$$f(t) = Ae^{i\phi(t)}.$$

Now we narrow the possibilities some more. Suppose f has the form $Ae^{i\phi(t)}$. Then

$$R(t_2 - t_1) = E[\xi\overline{\xi}]e^{i\phi(t_2) - i\phi(t_1)} = E[\xi\overline{\xi}]e^{i\phi(t_1+T) - i\phi(t_1)},$$

(where we wrote $t_2 - t_1 = T$) must be independent of t_1. This is enforced by the equation:

$$\frac{d}{dt}(\phi(t + T) - \phi(t)) = 0.$$

Then

$$\frac{d\phi}{dt}(t) = \frac{d\phi}{dt}(t + T) = \text{const} = \alpha.$$

Hence

$$\phi(t) = \alpha t + \beta$$

and f is of the form

$$f(t) = Ce^{i\alpha t}.$$

We have shown that the process $u(t, \omega) = \xi f(t)$ is stationary in the wide sense if $f(t) = Ce^{i\alpha t}$ and $E[\xi] = 0$ or if $f(t)$ is a constant.

4.2. Covariance and Spectrum

In the last section we presented an example of a stationary stochastic process in the wide sense, given by $u(t,\omega) = \xi e^{i\lambda t}$, where ξ is a random variable with mean zero. This stochastic process has a covariance of the form

$$R(T) = R(t_2, t_1) = R(t_2 - t_1) = E[|\xi|^2]e^{i\lambda T},$$

where $T = t_2 - t_1$. Now we want to generalize this example. First we try to construct a process of the form

$$u(t,\omega) = \xi_1 e^{i\lambda_1 t} + \xi_2 e^{i\lambda_2 t}.$$

Then $E[u] = E[\xi_1]e^{i\lambda_1 t} + E[\xi_2]e^{i\lambda_2 t}$ which is independent of t if $E[\xi_1] = E[\xi_2] = 0$. The covariance is

$$\begin{aligned} &E\left[(\xi_1 e^{i\lambda_1 t_2} + \xi_2 e^{i\lambda_2 t_2})(\overline{\xi_1} e^{-i\lambda_1 t_1} + \overline{\xi}_2 e^{-i\lambda_2 t_1})\right] \\ &\quad = E\left[|\xi_1|^2 e^{i\lambda_1 T} + |\xi_2|^2 e^{i\lambda_2 T} + \xi_1\overline{\xi}_2 e^{i\lambda_1 t_2 - i\lambda_2 t_1} + \overline{\xi}_1\xi_2 e^{i\lambda_2 t_2 - i\lambda_1 t_1}\right] \end{aligned}$$

which can be stationary only if $E[\xi_1\overline{\xi}_2] = 0$. Then $u(t,\omega)$ is stationary and

$$R(T) = E[|\xi_1|^2]e^{i\lambda_1 T} + E[|\xi_2|^2]e^{i\lambda_2 T}.$$

More generally, a process $u = \sum_j \xi_j e^{i\lambda_j t}$ is wide sense stationary if $E[\xi_j\overline{\xi_k}] = 0$ when $j \neq k$ and $E[\xi_i] = 0$. In this case

$$R(T) = \sum E\left[|\xi_j|^2\right] e^{i\lambda_j T}.$$

This expression can be rewritten in a more useful form as a Stieltjes integral. Recall that when g is a non-decreasing function of x the Stieltjes integral of a function f with respect to g is defined to be

$$\int f dg = \lim_{\max\{x_{i+1} - x_i\} \to 0} \sum f(x_i^*)[g(x_{i+1}) - g(x_i)]$$

where $x_i \leq x_i^* \leq x_{i+1}$. If g is differentiable then

$$\int_a^b f dg = \int_a^b f g' dx.$$

Suppose $g(x)$ is the step function:

$$g(x) = \begin{cases} 0 & x < c \\ q & x \geq c \end{cases},$$

with $a \leq c \leq b$. Then $\int_a^b f dg = f(c)q$. Now we define the function $F = F(k)$ by

$$F(k) = \sum_{\{j|\lambda_j \leq k\}} E[|\xi_j|^2]$$

i.e., $F(k)$ is the sum of the expected values of the squares of the amplitudes of the complex exponentials with frequencies less than k. $R(T)$ becomes

$$R(T) = \int_{-\infty}^{+\infty} e^{ikT} dF(k).$$

We shall now see that under some technical assumptions, this relation holds for all wide sense stationary stochastic processes. Indeed, we have:

THEOREM 4.1. *(**Khinchin**)*

(1) *If $R(T)$ is the covariance of a weakly defined wide sense stationary stochastic process such that*

$$\lim_{h \to 0} E\left[|u(t+h) - u(t)|^2\right] = 0$$

then $R(T) = \int e^{ikT} dF(k)$ for some non-decreasing function $F(k)$.

(2) *If a function $R(T)$ can be written as $\int e^{ikT} dF(k)$ for some non-decreasing function F, then there exists a weakly defined wide sense stationary stochastic process, satisfying the condition in part (1) of the theorem, that has $R(T)$ as its covariance.*

The proof of Khinchin's Theorem follows from (4.1). If $dF(k) = \phi(k)dk$, then $R(T) = \int e^{ikT}\phi(k)dk$ and $\phi(k)$ is called the spectral density of the process. Thus Khinchin's theorem states that the covariance function is a Fourier transform of the spectral density. Hence, if we know $R(T)$ we can compute the spectral density by

$$\phi(k) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ikT} R(T) dT.$$

EXAMPLE. In the case of white noise we have $R(T) = \delta(T)$. Its spectral density is $\phi(k) = 1/2\pi$, thus all frequencies have the same amplitude. The adjective "white" comes from the fact that in white light all frequencies are present with the same amplitude. Any random signal which is not white noise is called colored noise.

4.3. Application: The Inertial Spectrum of Turbulence

.

To illustrate these constructions, we now derive the "inertial range" spectrum of fully developed turbulence. The equations of motion will not be written down because they will not be used directly.

Consider turbulence far from walls, with the Reynolds number $Re = UL/\nu$ very large, where U is a typical velocity in the flow, L is a length scale for the turbulence, and ν is the viscosity; the dimensionless number Re measures the amount by which the "inertial ", i.e., nonlinear, terms in the equations of motion dominate the viscous dissipation, and is large in fully developed turbulence. The movement of energy from scale to scale, i.e., from one k to another, is described by the nonlinear terms in the equation of motion. The flow is driven by large scale forcing (for example, in the case of meteorolgy, by the rotation of the earth around its axis and around the sun); one assumes that by the time the energy has moved to large wave numbers k (i.e., small wavelengths) the geometry of the forcing has been forgotten and the flow can be viewed as approximately homogeneous (translation invariant) and isotropic, (rotation inavariant) and its spectral properties are universal (i.e., independent of specific geometry and forcing).

The velocity field in three space dimensions is a vector quantity: $\mathbf{u} = (u_1, u_2, u_3)$. One can define a correlation tensor

$$R_{ij}(\mathbf{r}) = E[u_i(\mathbf{x} + \mathbf{r})u_j(\mathbf{x})],$$

and find

$$R_{ij}(\mathbf{r}) = \int_{-\infty}^{\infty} dF_{ij}(\mathbf{k}),$$

where $\mathbf{k} = (k_1, k_2, k_3)$. Without loss of generality in what follows one can write $dF_{ij}(\mathbf{k}) = \Psi_{ij}(\mathbf{k})dk_1dk_2dk_3$ (this is so because all we will care about is the dimensions of the various quantities, which is not affected by their possible lack of smoothness). Finally, one can define

$$E(k) = \int_{k_1^2+k_2^2+k_3^2=k^2} (\Psi_{11} + \Psi_{22} + \Psi_{33})\, dA(\mathbf{k}).$$

$E(k)$ is the average of the trace of the tensor $\Psi_{ij}(\mathbf{k})$ over the sphere of radius k in wave number space and is a function of $k = \sqrt{k_1^2 + k_2^2 + k_3^2}$ only. One can see from the various identities that have been derived that $E[u^2] = \int_0^\infty E(k)dk$, where $u^2 = u_1^2+u_2^2+u_3^2$. $E(k)$ is the spectrum of the flow.

The energy is proportional to the square of the velocity while energy dissipation, modelled in the equations of motion by the Laplace

operator acting on the velocity, is proportional to the square of the derivatives of the velocity; in spectral variables, (i.e., after Fourier transformation), the energy is proportional to the square of $\hat{u}(k)$ while the dissipation is proportional to the square of $k\hat{u}(k)$, where $\hat{u}$ is the (random) Fourier transform of the velocity $u = u(x)$. It is plausible that when Re is large the energy resides in a range of k's disjoint from the range of k's where the dissipation is taking place, and indeed experimental data show it to be so; specifically, there exist wave numbers k_1, k_2 such that

$$\int_0^{k_1} E(k)dk \sim \int_0^{\infty} E(k)dk, \quad \int_{k_2}^{\infty} k^2 E(k)dk \sim \int_0^{\infty} k^2 E(k)dk,$$

with $k_1 \ll k_2$. The range of k's such that $k_1 < k < k_2$ is the "inertial range" of wave numbers; the name is a bit of a misnomer because it implies that in that range the mechanics is purely "inertial", free of viscosity effects, but we shall see that this is not so. This is the range of wave numbers k we now focus on.

We will be relying on dimensional analysis. Suppose a variable a is a function of variables $a_1, a_2, \ldots, a_m$, $b_1, b_2, \ldots, b_k$, where $a_1, \ldots, a_m$ have independent units. For example a_1 could have the dimension of length measured in units L (say kilometers) and a_2 could be a time, measured in units T of time (say seconds), while the units of $b_1, \ldots, b_k$, can be formed from the units of $a_1, a_2, \ldots, a_m$; in the example just used, b_1 could be a velocity, whose units are L/T. Then there exist dimensionless variables

$$\Pi = \frac{a}{a_1^{\alpha_1} \cdots a_m^{\alpha_m}}, \quad \Pi_i = \frac{b_i}{a_1^{\alpha_{i1}} \cdots a_m^{\alpha_{im}}}, \quad i = 1, \ldots, k,$$

where the α_i, α_{ij} are simple fractions, such that Π is a function of the Π_i:

$$\Pi = \Phi(\Pi_1, \ldots, \Pi_k).$$

This is just a consequence of the requirement that a physical relationship be independent of the size of the units of measurement. At this stage nothing can be said about the function Φ. Now suppose the variables Π_i are small or large, (the two cases are indistinguishable, because an unknown function of x is also an unknown function of $1/x$), and assume that the function Φ has a non-zero finite limit as its arguments tend to zero or to infinity; then $\Pi \sim$ constant, and one finds a power monomial relation between a and the a_i: $a = a_1^{\alpha_1} \cdots a_m^{\alpha_m}$. This is a "complete similarity relation. If the function Φ does not have the assumed limit, it may happen that for Π_1 small or large, $\Phi(\Pi_1) = \Pi_1^{\alpha} \Phi_1(\Pi_1) + \ldots$, where the dots denote lower order terms, α

is a constant, the other arguments of Φ have been omitted and Φ_1 has a finite non-zero limit. One can then obtain a monomial expression for a in terms of the a_i and b_i, with undetermined powers which must be found by means other than dimensional analysis. The resulting power relation is an *incomplete* similarity relation. The exponent α is known in the physics literature as an anomalous scaling exponent; in physics incomplete similarity is usually discussed in the context of the renormalization group, see Chapter 5. Of course one may well have functions Φ with neither kind of similarity.

We now apply these scaling ideas to the spectrum. The spectrum in the inertial range $E(k)$ is a function of k, of the viscosity ν, of the length scale L, of the amplitude U of the typical velocity in the flow, and of the rate of energy dissipation ϵ. That last variable belongs here because energy is transferred from the low k domain through the inertial range into the large k domain where it is dissipated; the fact that ϵ belongs in the list was the brilliant insight of Kolmogorov.

Our basic units are the units of length L and of time T. The units of the viscosity are L^2/T, those of ϵ L^2/T^3, those of k L, while the identity $E[u^2] = \int E(k)dk$ show that the units of E are L^3/T^2. Dimensional analysis yields $E(k)/(\epsilon^{-2/3}k^{5/3}) = \Phi(Re, Lk)$ for some unknown function Φ of the two large arguments Re and Lk; Re is large because this is the condition for fully developed turbulence to appear and Lk is large in the inertial range of scales. If the function Φ has a finite non-zero limit C as its arguments grow one can deduce $E(k) = C\epsilon^{2/3}k^{-5/3}$—the famous Kolmogorov-Obukhov scaling law for the inertial range of fully developed turbulence, the cornerstone of turbulence theory.

This law is not fully satisfactory, for various reasons, and various correction schemes have been proposed over the years. In recent years it has been shown that it is mathematically satisfactory, as well as in agreement with experiment, to set in the relation above

$$\Phi(Re, Lk) = C(Re)\epsilon^{2/3}k^{-5/3}(Lk)^{d/ln(Re)},$$

where $C(Re)$ is a function of Re and d is a positive constant.

4.4. Random Measures

We now embark upon a development which is a little bit outside of the main line of this course; it has wide applications but will not be used further in these notes. In particular, we will show that arbitrary wide-sense stationary processes can be represented as convolutions of non-random functions with certain simple processes (often Brownian motion). An important special case of this representation is the stochastic Fourier transform, which exists whenever the covariance

function exists, but does not require that the process itself have samples to which the standard Fourier transform can be applied; this is a key building block in the study of turbulence, signal processing, and quantum theory.

Given a probability space $(\Omega, \mathcal{B}, P)$ consider the set of random variables $f(\omega)$, where ω is a random variable, such that $E[f\bar{f}] < \infty$. We refer to this set as $L_2(\Omega, \mathcal{B}, P)$. We now construct a one-to-one mapping $L_2(\Omega, \mathcal{B}, P) \to L_2(A, \mu)$, where A is a subset of the t-axis and μ is a measure on A. Consider $\mathcal{A}$, an algebra of subsets of A, given by

$$\mathcal{A} = \left\{ A_i \subset A \mid C(A_i) \in \mathcal{A} \text{ and } \bigcap_{i=1}^{n} A_i, \bigcup_{i=1}^{n} A_i \in \mathcal{A} \right\}$$

where n is finite. (An algebra is much like a σ-algebra, with the exception that we do not require that the union of a countably infinite family of subsets belong to the algebra, a detail which is important to a rigorous analysis, but which we will omit here.)

Now consider the triple $(A, \mathcal{A}, \mu)$ where μ is a rule which to each subset $A_i \in \mathcal{A}$ assigns a number such that

(1) $\mu(A_i) \geq 0$.
(2) $\mu(A) = 1$.
(3) $\mu(\emptyset) = 0$.
(4) $A_i \cap A_j = \emptyset \Rightarrow \mu(A_i \cup A_j) = \mu(A_i) + \mu(A_j)$.

(Again, note that we are concerned only with finitely many A_i.) Next construct a random variable $\rho = \rho(A_i, \omega)$ where $A_i \in \mathcal{A}$ and $\omega \in \Omega$ (remember that a random variable is a function defined on Ω) which has the following properties:

(1) $A_i \cap A_j = \emptyset \Rightarrow \rho(A_i \cup A_j, \omega) = \rho(A_i, \omega) + \rho(A_j, \omega)$.
(2) $\rho(A_i, \omega)$ is square integrable, i.e., $E[\rho(A_i, \omega)\bar{\rho}(A_i, \omega)] < \infty$.
(3) $\rho(\emptyset, \omega) = 0$.
(4) $A_i, A_j \subset A \Rightarrow E[\rho(A_i, \omega)\bar{\rho}(A_j, \omega)] = \mu(A_i \cap A_j)$.

Note that the properties listed above imply that $\mu(A) \geq 0$, since then

$$\mu(A) = \mu(A \cap A) = E[\rho(A, \omega)\bar{\rho}(A, \omega)] \geq 0.$$

μ is called the structure function of ρ. Just as a stochastic process is a function of both ω and t, so is a random measure a function of both ω and the subsets A_i of A.

Now define $\chi_{A_i} = \chi_{A_i}(t)$, the characteristic function of the subset A_i of the t-axis, to be

$$\chi_{A_i} = \begin{cases} 1 & t \in A_i \\ 0 & \text{otherwise} \end{cases},$$

and consider a function $q(t)$ of the form

$$q(t) = \sum c_i \chi_{A_i}(t).$$

We consider the case where $\{A_i\}$ is a finite partition of A, i.e., there are only finitely many A_i, $A_i \cap A_j = \emptyset$ for $i \neq j$, and $\bigcup A_i = A$. Thus $q(t)$ takes on only a finite number of values. To this function $q(t)$ assign the random variable

$$f(\omega) = \sum c_i \rho(A_i, \omega).$$

Hence each characteristic function of a subset is replaced by the random variable which the random measure assigns to the same subset; thus this substitution transforms a function of t into a function of ω, i.e., into a random variable.

Now consider the product $q_1(t)\overline{q_2}(t)$ of two functions of the form

$$q_1 = \sum_{j=1}^{n} c_j \chi_{A_j}(t), \quad q_2 = \sum_{k=1}^{m} d_k \chi_{B_k}(t)$$

where the $\{B_i\}$ is another finite partition of A. It is not necessary for n and m to be equal. There is a finite number of intersections of the A_i's and B_i's and on each of these subsets the product from $q_1 q_2 = \sum_{j=1}^{n} c_j \chi_j(A_j) \times \sum_{k=1}^{m} d_k \chi_k(B_k)$,

$$q_1 \overline{q_2} = \left(\sum_{j=1}^{n} c_j \chi_j(A_j) \right) \left(\sum_{k=1}^{m} \overline{d_k} \chi_k(B_k) \right)$$

takes on a constant value $c_j \overline{d_k}$. Thus the same construction allows us to assign a random variable $f_1 \overline{f_2}$ to the product $q_1 \overline{q_2}$. Since

$$f_1(\omega) = \sum c_j \rho(A_j, \omega), \quad f_2(\omega) = \sum d_k \rho(B_k, \omega),$$

we conclude that

$$\begin{aligned} E[f_1 \overline{f_2}] &= E\left[\sum_{j=1}^{n} \sum_{k=1}^{m} c_j \overline{d_k} \rho(A_j, \omega) \rho(B_k, \omega) \right] \qquad (4.2) \\ &= \sum_{j=1}^{n} \sum_{k=1}^{m} c_j \overline{d_k} E\left[\rho(A_j, \omega) \rho(B_k, \omega) \right] \\ &= \sum_{j=1}^{n} \sum_{k=1}^{m} c_j \overline{d_k} \mu(A_j \cap B_k) \\ &= \int q_1 \overline{q_2} \mu(dt). \end{aligned}$$

Thus we have established a mapping between random variables with finite mean squares and functions of time with finite square integrals (i.e., between the random variables $f(\omega)$ and functions $q(t)$ such that $\int q_1(t)\overline{q_2}(t)\mu(dt)$ is finite.) Although we have defined the mapping only for functions $q(t) = \sum c_i \chi_{A_i}(t)$, an argument that we omit enables us to extend the mapping to all random variables and functions of t with the square integrability properties listed above.

Example. We now show in detail how this construction works for a very special case. Say we are given a probability space (Ω, B, P) and three subsets of the t axis: $A_1 = [0,1), A_2 = [1,3)$, and $A_3 = [3, 3\frac{1}{2}]$. Each A_i is assigned a random variable $\rho_i(\omega) = \rho(A_i, \omega)$ which has mean 0 and variance equal to the length of A_i. For example, $\rho_1(\omega)$ has mean 0 and variance 1, etc. The variables ρ_1, ρ_2, ρ_3 are independent, and $E[\rho_i \rho_j] = 0$ for $i \neq j$ where $E[\rho_i^2]$ is the length of the i^{th} interval. Moreover,

$$\chi_1 = \begin{cases} 1, & 0 \leq t < 1 \\ 0, & \text{elsewhere} \end{cases},$$

$$\chi_2 = \begin{cases} 1, & 1 \leq t < 3 \\ 0, & \text{elsewhere} \end{cases},$$

and

$$\chi_3 = \begin{cases} 1, & 3 \leq t \leq 3\frac{1}{2} \\ 0, & \text{elsewhere} \end{cases}$$

where $\int \chi_i \chi_j dt = 0$ for $i \neq j$ and $\int \chi_i^2 dt$ is the length of the i^{th} interval.

Now take a function of the form $q_1(t) = \sum_i c_i \chi_i(t)$ where the c_i's are constants. Then

$$q_1(t) \to f_1(\omega) = \sum_{i=1}^{3} c_i \rho_i(\omega).$$

Suppose we have another function $q_2(t)$

$$q_2(t) = \sum_{j=1}^{3} d_j \chi_j(t) \to f_2(\omega) = \sum_{j=1}^{3} d_j \rho_j(\omega).$$

Then

$$E[f_1 \overline{f_2}] = E\left[\sum_{i=1}^{3}\sum_{j=1}^{3} c_i \overline{d_j} \rho_i \rho_j\right] = \sum_{j=1}^{3} c_j \overline{d_j} E\left[\rho_j^2\right] = \sum_{j=1}^{3} c_j \overline{d_j} \mu(A_j)$$

where $\mu(A_j)$ is the length of A_j. Notice also that

$$\int_0^{3\frac{1}{2}} q_1(t)q_2(t)dt = \int_0^{3\frac{1}{2}} \sum_{i=1}^{3}\sum_{j=1}^{3} c_i\overline{d_j}\chi_i(t)\chi_j(t) = \sum_j c_j\overline{d_j}\mu(A_j)$$

which verifies that $q(t) \to f(t)$ so $E[f_1\overline{f_2}] = \int q_1(t)q_2(t)\mu dt$ as in (4.2).

Now approximate every square-integrable function on A (i.e., such that $\int_A q\bar{q}\,d\mu$ is finite) by a step function, construct the corresponding random variable, and take the limit, as the approximation improves, of the sequence of random variables obtained in this way. This makes for a mapping of square integrable functions on A onto random variables with finite mean squares. This mapping can be written as

$$f(\omega) = \int q(s)\rho(ds,\omega)$$

where the variable t has been replaced by s for convenience. Now view a stochastic process u as a family of random variables labelled by the parameter t (i.e., there is a random variable u for every value of t) and apply the representation just derived at each value of t, so

$$u(t,\omega) = \int q(t,s)\rho(ds,\omega).$$

Assume $u(t,\omega)$ is wide-sense stationary. Then the covariance of u is

$$\begin{aligned} R(t_2-t_1) &= E[u(t_1,\omega)\overline{u_2}(t_2,\omega)] \\ &= E\left[\int q(t_1,s)\rho(ds_1)\int \bar{q}(t_2,s_2)\rho(ds_2)\right] \\ &= E\left[\int q(t_1,s_1)\bar{q}(t_2,s_2)\rho(ds_1)\bar{\rho}(ds_2)\right] \\ &= \int q(t_1,s_1)\bar{q}(t_2,s_2)E[\rho(ds_1)\bar{\rho}(ds_2)] \\ &= \int q(t_1,s)\bar{q}(t_2,s)\mu(ds). \end{aligned}$$

One can show that the converse is also true: if the above holds then $u(t,\omega) = \int q(t,s)\rho(ds,\omega)$ with $E[\rho(ds)\bar{\rho}(ds)] = \mu(ds)$. Note that in all of the above equality holds in a mean-square (L_2) sense and thus little can be said about the higher moments.

EXAMPLE. If $u = u(t,\omega)$ is a wide-sense stochastic process then

$$R(t_1,t_2) = \int e^{ik(t_2-t_1)}dF(k).$$

Let $q(t_2, k) = e^{ikt_2}$ and $\bar{q}(t_1, k) = e^{-ikt_1}$ (taking $k = s$). Then

$$u(t, \omega) = \int e^{ikt} \rho(dk, \omega) \tag{4.3}$$

where $E[\bar{\rho}(dk)] = dF(k)$. We have just shown that $dF(k)$ is the energy density in the interval dk. The Fourier transform of u does not exist in the usual sense (i.e., $\int u(t, \omega) e^{ikt} dt$ does not exist), but for (4.3) it is sufficient for $E[|u(t)|^2]$ to exist for each t.

EXAMPLE. Suppose $dF(k) = \phi(k)dk$. Then

$$\int e^{ik(t_2 - t_1)} dF(k) = \int e^{ikt_2} \sqrt{\phi(k)} e^{-ikt_1} \sqrt{\phi(k)}\, dk.$$

Recall that $\phi(k) \geq 0$. Write $\sqrt{\phi(k)} = \hat{h}(k) = \widehat{h(t)}$, where $h(t)$ is the inverse Fourier transform of $\hat{h}(k)$. Recall that for a function $h(t)$ we have:

$$\widehat{h(t - t_1)} = e^{-ikt_1} \hat{h}(k),$$

so the covariance $R(t_1, t_2)$ can be written

$$R(t_1, t_2) = \int \overline{\widehat{h(t - t_2)}} \widehat{h(t - t_1)} dk.$$

Since the Fourier transform preserves the inner products we have

$$R(t_1, t_2) = \int \overline{h(t - t_2)} h(t - t_1) dt,$$

and by changing t to s we obtain

$$R(t_1, t_2) = \int \overline{h(s - t_2)} h(s - t_1) \mu(ds),$$

where $\mu(ds) = ds$. Applying our representation, we get $u(t, \omega) = \int h(s - t) \rho(ds)$ where $E[|\rho(ds)|^2] = ds$. Note that the random measure constructed as increments of Brownian motion at instants ds apart has this property. Thus, any wide-sense stationary stochastic process with $dF(k) = \phi(k)dk$ can be approximated as a sum of translates (in time) of a fixed function, each translate multiplied by independent Gaussian random variables. This is the "moving average" representation.

4.5. Prediction for Stationary Stochastic Processes

Consider a stationary stochastic process $u(t, \omega)$ and suppose we are given the values $U_1, U_2, \ldots, U_n$ of $u(t, \omega)$ for $s \leq t$. (For clarity we denote by capital letters values that are known.) The question we will address in this section is how to predict a value for $u(t + T, \omega)$ based

on the information given. For simplicity we shall do so only for a stationary random sequence.

DEFINITION. A stationary random sequence is a collection $u(t,\omega)$ of random variables for $t = 1, 2, 3, \ldots$ as well as for $t = -1, -2, -3, \ldots$ such that the joint distribution of any subset is known, subject to the obvious compatibility conditions, and such that all the distributions are invariant under the transformation $t \to t + T$ for T integer.

Assume $E[u(t)] = 0$. The covariance

$$R(T) = E[u(t+T)\overline{u(t)}]$$

where $T \in \mathbb{Z}$ satisfies, as before:

(1) $R(0) \geq 0$.
(2) $|R(T)| \leq R(0)$.
(3) $R(-T) = \overline{R(T)}$.
(4) $\sum_{j,l} R(j-l)a_j\overline{a_l} \geq 0$.

If $u(t,\omega) = \xi(\omega)f(t)$ is stationary, then $f(t) = Ce^{ikt}$, with t integer. But note that since t is an integer, $e^{i(k+2\pi)t} = e^{ikt}$, so we can assume $0 \leq k \leq 2\pi$. The covariance then has the spectral representation

$$R(T) = \int_{-\pi}^{\pi} e^{ikT} dF(k).$$

For simplicity we assume that $f = F'(k)$ exists so that

$$R(T) = \int_{-\pi}^{\pi} e^{ikT} f(k) dk.$$

Recall that if $f(k)$ is a periodic function with period 2π then its Fourier transform is

$$f(k) = \sum_{n=-\infty}^{\infty} a_n e^{ink},$$

where

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(k) e^{-ink} dk.$$

Note that $R(-n)$ is the n^{th} Fourier component of $f(k)$ and

$$f(k) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} R(n) e^{-ink}.$$

(The factor $1/2\pi$ is broken up differently from what was done earlier for convenience.)

Consider the problem of finding a good estimate for $u(t+m,\omega)$ when we have values $U(t-n)$, $U(t-(n-1))$, ..., $U(t-1)$. We would like to find a random variable $\tilde{u}(t+m,\omega)$ such that

$$E\left[|u(t+m,\omega)-\tilde{u}(t+m,\omega)|^2\right]$$

is as small as possible. We know from earlier work that

$$\tilde{u}(t+m,\omega)=E[u(t+m,\omega)|u(t-1),u(t-2),\ldots,u(t-n)].$$

The way to evaluate $\tilde{u}$ is to find a basis $\{\phi_i\}$ in the space Q of functions of $\{u(t-n),\ldots,u(t-1)\}$, expand $\tilde{u}$ in this basis, i.e.,

$$\tilde{u}=\sum_{j=1}^{n} a_j\phi_j(u(t-1),\ldots,u(t-n)),$$

and calculate the coefficients a_j of the expansion. This is hard in general. We simplify the problem by looking only for the best approximation in the span of $\{u(t-1),\ldots,u(t-n)\}$, i.e., we look for a random variable

$$\tilde{u}(t,\omega)=\sum_{j=1}^{n} a_j u(t-j,\omega).$$

This is called linear prediction. Note that the span of $u(t-j,\omega)$, call it L, is a closed linear space, and therefore the best linear prediction minimizes

$$E\left[|u(t+m,\omega)-\tilde{u}(t+m,\omega)|^2\right]$$

for $\tilde{u}$ in L. What we have to do is to find $\{a_j\}_{j=1}^n$, such that

$$E\left[\left|u(t+m,\omega)-\sum_{j=1}^{n} a_j u(t-j,\omega)\right|^2\right]$$

is as small as possible. We have

$$E\left[|u-\tilde{u}|^2\right] =$$
$$E\left[\left(u(t+m)-\sum_j a_j u(t-j)\right)\overline{\left(u(t+m)-\sum_l a_l u(t-l)\right)}\right]$$
$$= E\left[u(t+m)\overline{u(t+m)} - \sum_l \overline{a_l} u(t+m)\overline{u(t-l)}\right.$$
$$\left. - \sum_j a_j \overline{u(t+m)} u(t-j) + \sum_j \sum_l a_j \overline{a_l} u(t-j)\overline{u(t-l)}\right]$$
$$= R(0) - 2Re\left(\sum_j \overline{a_j} R(m+j)\right) + \sum_j \sum_l a_j \overline{a_l} R(l-j),$$

which is minimized when

$$\frac{\partial E\left[|u-\tilde{u}|^2\right]}{\partial \overline{a_j}} = -R(m+j) + \sum_{l=1}^{n} a_l R(j-l) = 0 \tag{4.4}$$

for $j = 1, \ldots, n$. The uniqueness of the solution of the system (4.4) and the fact that this procedure gives a minimum are guaranteed by the orthogonal projection theorem for closed linear spaces (see Section 1.1).

Rewrite (4.4) in terms of the Fourier transform. The spectral representation of $R(T)$ is

$$R(T) = \int_{-\pi}^{\pi} e^{ikT} dF(k).$$

Assume that $dF(k) = f(k)dk$. Then (4.4) becomes

$$\int_{-\pi}^{\pi} \left(-e^{i(j+m)k} + \sum_{l=1}^{n} a_l e^{i(j-l)k}\right) f(k)dk = 0.$$

Putting e^{ijk} outside the parentheses we get

$$\int_{-\pi}^{\pi} e^{ijk} \left(e^{imk} - \sum_{l=1}^{n} a_l e^{-ilk}\right) f(k)dk = 0. \tag{4.5}$$

We now solve this equation using complex variables. Let $g(z)$ be a complex valued function of a complex variable $z = x + iy$ and C be a closed curve bounding a simply connected region.

(1) If $g(z)$ is analytic on and inside C then $\oint_C g(z)dz = 0$.
(2) If g is an analytic function defined on C then there exists a function f^* analytic on and inside C such that $f = f^*$ on C.

(3) Suppose g is analytic outside C and $g(\infty) = 0$. Then g can be written in the form $g(z) = \sum a_n z^{-n}$.
(4) If g is bounded and analytic everywhere then g is constant.

Return to formula (4.5). Our aim is to determine the set of coefficients a_l, $l = 1, \ldots, n$, which will give us the best linear approximation $\tilde{u}(t+m, \omega)$ to $u(t+m, \omega)$. On the unit circle $|z| = 1$ z may be written as $z = e^{ik}$ where $-\pi \leq k \leq \pi$. So the function

$$\Phi(k) = \sum_{l=1}^{n} a_l e^{-ikl}$$

can be thought to be an analytic complex function defined on the unit circle. Suppose $f(k)$ is smooth enough on the unit circle. Let $\Phi^*(z)$ and $f^*(z)$ be the extensions of the functions Φ and f to the complex plane,

$$\Phi^*(z) = \frac{a_1}{z} + \frac{a_2}{z^2} + \ldots + \frac{a_n}{z^n}.$$

Note that $\Phi^*(\infty) = 0$. Then (4.5) can be written as

$$\int_{|z|=1} z^j \left(z^m - \Phi^*(z)\right) f^*(z) dz = 0.$$

In the case $n = \infty$ we seek $\Phi^*(z)$ that satisfies:

(1) $\Phi^*(z)$ is convergent on the unit circle.
(2) $\Phi^*(\infty) = 0$.
(3) $\Phi^*(z)$ is analytic outside the unit circle.
(4) $(z^m - \Phi^*(z)) f^*(z)$ is analytic on and inside the unit circle (i.e, it may be written as a Taylor series there).

If we find such a function Φ^*, the coefficient a_l in its expansion will satisfy (4.5).

Consider an example. Suppose that we are dealing with a stationary random sequence whose covariance is

$$R(T) = \begin{cases} Ca^T, & T > 0 \\ Ca^{-T}, & T < 0 \end{cases},$$

where $C > 0$ and $0 < a < 1$. We have seen that the covariance of any stationary process must satisfy the conditions:

(1) $R(0) > 0$.
(2) $|R(T)| \leq R(0)$.
(3) $R(-T) = R(T)$.
(4) $\sum_{i,j} R(i-j) z_i \overline{z_j} \geq 0$ for any $z \in \mathbb{C}$ (R is positive definite).

It is first necessary to check that the problem makes sense, that is, that there exists a stationay process which has the given covariance—not every function is the covariance of some stationary process. We have also seen that condition (4) holds if R is the Fourier transform of a non-negative function. To check this we just calculate the Fourier transform f of R:

$$f(k) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} e^{-ink}R(n) = \frac{C}{2\pi}\sum_{n=1}^{\infty} e^{ink}R(n) + \frac{C}{2\pi} + \frac{C}{2\pi}\sum_{n=1}^{\infty} e^{-ink}R(n).$$

Hence

$$f(k) = \frac{C}{2\pi}(1-a^2)\frac{1}{(e^{ik}-a)(e^{-ik}-a)} > 0.$$

Therefore there exists a stationary stochastic sequence such that this R is its covariance.

The extension of $f(k)$ to the complex plane is:

$$f^*(z) = C_1\frac{1}{(z-a)(z^{-1}-a)} = \frac{C_1 z}{(z-a)(1-az)},$$

where $C_1 = (C/2\pi)(1-a^2)$. As was shown above, the function

$$(z^m - \Phi^*(z))\frac{z}{(z-a)(1-az)}$$

has to be analytic inside the unit circle.

Consider the special case $m = 1$. To make the function

$$(z - \Phi^*(z))\frac{z}{(z-a)(1-az)}$$

analytic inside the unit circle we should choose $\Phi^*(z)$ such that

(1) $\Phi^*(a) = a$ and
(2) $\Phi^*(z) = z^{-1}\Psi(z)$, where $\Psi(z)$ is analytic inside the unit circle.

It also follows from the previous discussion that $\Psi(z)$ must be analytic outside the unit circle and vanish at infinity. Consequently $\Psi(z)$ must be constant. To satisfy (1) Ψ must equal a^2. Thus

$$\Phi^*(z) = \frac{a^2}{z}.$$

And hence the best approximation of $u(t+1,\omega)$ is

$$u(t+1,\omega) = a^2 u(t,\omega).$$

The best guess of the next value of the sequence is a^2 times the last value, where a is the factor by which the covariance decreases over a unit time interval.

4.6. Data Assimilation

We now turn to the topic of data assimilation, which could have been discussed at the end of Chapter 3 but which has been set here so that the presentation can be read along with the related discussion of prediction for stationary processes in the previous section.

There are many situations where one wants to make predictions on the basis of models which are not accurate enough, but which can be supplemented by current data. The canonical example is meteorology, where at any one time one has an incomplete description of the current weather, the equations of motion provide an incomplete description of the atmosphere, but data are coming in all the time. The use of data together with a model to assess the current state of a system and/or to make predictions is called "data assimilation", and the algorithms for doing that are called "filters."

A useful model of a situation where data assimilation is needed consists of a stochastic differential equation

$$d(\mathbf{x}) = \mathbf{f}(\mathbf{x}, t)d\mathbf{x} + g(\mathbf{x}, t)d\mathbf{w}, \tag{4.6}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is an n-dimensional vector, $d\mathbf{w}$ is an n-dimensional BM, f is an n-dimensional vector function, and g is a scalar (i.e., an n by n diagonal matrix of the form gI, where g is a scalar and I is the identity matrix). The BM encapsulates all that is not known in this model. The initial data $\mathbf{x}(0)$ are assumed given and they may be random as well.

As the experiment unfolds it is observed, and the values $\mathbf{y}_i$ of a "measurement process" are recorded at times $t_i = i\delta$, where δ is a fixed time interval; they are related to the evolving "state" $\mathbf{x}(t)$ by

$$\mathbf{y}_i = \mathbf{h}(\mathbf{x}_i, t_i) + GW_i, \tag{4.7}$$

where $\mathbf{h}$ is a k-dimensional vector, where in principle $k \leq n$ (but in what follows we assume $k = n$), $\mathbf{h}$ is a nonlinear function, G is a diagonal matrix with non-zero diagonal terms, $\mathbf{x}_i = \mathbf{x}(i\delta)$, and W_i is a vector whose components are independent Gaussian variables of mean zero and variance 1, independent also of the BM's in the equation. Equation (4.7) says that the observations are noisy, with Gaussian noise. Now the problem is to estimate $\mathbf{x}$ on the basis of equation (4.6) and the observations $\mathbf{y}_i$. We are interested in cases where simply rerunning the problem with a different sample of BM will not do because the different samples differ more than we can tolerate; the observations should narrow down the range of possible $\mathbf{x}(t)$. The solution of the problem is of course the process $\hat{\mathbf{x}} = E[\mathbf{x}(t)\bar{\mathbf{y}}(t)]$, where $\bar{\mathbf{y}}(t)$ is the

sequence $\mathbf{y}_1, \mathbf{y}_2, \dots$ for indices j such that $j\delta \le t$. This, as we know, is the best approximation of $\mathbf{x}(t)$ given $\bar{\mathbf{y}}$.

If the system (4.6) is linear and the data are Gaussian (or, as a special case, not random) then the solution of the problem is a Gaussian process. Its means and covariances can be calculated from those of the various functions that appear in the equation. This is the "Kalman filter", or "Kalman-Bucy filter", a mainstay of engineering. It provides a non-random solution of a random problem. This is not the place to present the algebra involved in deriving the full Kalman filter procedure, and we are content with a simple example as an illustration.

Suppose equation (4.6) is scalar and for the form reduces to $dx = 0$, i.e., the data $x(0)$, which we take as Gaussian with mean 0 and variance σ, do not change in time, $x(t) = x(0)$ (we write x rather than $\mathbf{x}$ for a scalar variable). Suppose the observation process is $y_i = x_i + gW_i$, with the W_i indepedent Gaussian variables of mean zero and variance 1. The variance of y_1 is $\sigma + g^2$, the projection of x_1 on y_1 is $y_1\sigma/(\sigma + g^2) = \hat{x}$, the filtered estimate. Note the following: If the variance g of the observation noise is large the observation adds little to the accuracy of the simplest estimate unaided by observations in which the variable x_1 is estimated by its mean; on the other hand if the variance of the observation noise is small, the observation is accurate, and the estimate reduces to equating the estimate to the observation. Thus the filter neatly blends in the information from the "equation" and the information from the observation, weighing their reliability as measured by the variances of the noises.

Now consider the general nonlinear case of equation (4.6). We have to estimate the variables $\mathbf{x}(t)$ and the natural thing to do is try to evaluate their probability density as it evolves in time. The initial datum $\mathbf{x}$ is known and so is its probability density, so all we have to do is evaluate sequentially the density of $\mathbf{x}_{i+1}$ assuming that we know the density of $\mathbf{x}_i$.

Let P_i be the probability density of $\mathbf{x}$ at time $i\delta$ (taking into account the observations before that time and at that time). To find the probability density of $\mathbf{x}$ at times $i\delta < t < (i+1)\delta$ (before any more observations come in) one can sample the density P_i, evolve the samples independently by (4.6), and whenever needed reconstruct a density by, for example, parametric estimation (see Section 2.5). The challenge is to modify the density at time $(i+1)\delta$, when new data must be taken into account. This can be done by Bayesian estimation (see Chapter 2).

Bayes theorem says that

$$P(\mathbf{x}|\bar{\mathbf{y}}_{i+1}) = \frac{P(\mathbf{y}_{i+1}|\mathbf{x}(t))P(\mathbf{x}|\bar{\mathbf{y}}_i)}{\int P(\mathbf{y}_{i+1}|\mathbf{x})P(\mathbf{x}|\bar{\mathbf{y}}_i)d\mathbf{x}}. \tag{4.8}$$

Where $P(\mathbf{x}|\bar{\mathbf{y}}_n)$ is the probability density determined from equation (4.6) taking into account the data up to and including time $i\delta$ but not the data at $(i+1)\delta$, $P(\mathbf{y}_{i+1}|\mathbf{x}(t))$ is the probability of finding the data if one knows the value $\mathbf{x}((i+1)\delta$ of the unknown vector $\mathbf{x}$, and the integral in the denominator is what is needed to normalize the probabilities. The connection with Bayesian estimation is made by taking $P(\mathbf{x}|\bar{\mathbf{y}}_i)$, the density in which the new data have not yet been taken into account, as the prior density, and then taking the density after the data have been used as the posterior density.

In words, formula (4.8) says that the new density given the new data is the product of the probability of getting the data if the values of the samples of the distribution were known, multiplied by the prior probability for samples of the distribution, the whole thing properly normalized. The probability of getting the data if the values of the samples were known can be obtained from the observation equation (4.7):

$$P(y_i \le x_i < y_i + dy_i) = \frac{\exp\left(s_i - \dfrac{h(\mathbf{x},t)}{g_{ii}}\right)}{\sqrt{\pi g_{ii}}}\, ds_i, \tag{4.9}$$

where g_{ii} is a diagonal entry of the matrix G. The formula (4.8) can be evaluated as follows: We can find n samples of P_i, and evolve them by equation (4.6). The density that one can in principle reconstruct from the positions of these samples after evolution does not take into account the new information at time $(i+1)\delta$, and we will use it as the prior density at the new time. The new information (i.e., the observation) at the time $(i+1)\delta$ makes it possible to assign a probability to each new sample position; if the first sample, for example, is at position $\mathbf{x}_1$; then its probability is given by formula (4.9).

Before taking the data into accounts assign to each sample a weight, say the weight 1. Take the data into account by replacing these uniform weights by $Z^{-1}p_i^{\text{new}}$, where the p_i^{new} come from formula (4.9)—they take into account the fact that the new observations make some sample positions that are far from the observation unlikely and those that are near the observation more likely. Choose Z so that the sum of the new weights is 1. The sample positions in $\mathbf{x}$-space are unchanged. We now have samples whose positions have been determined by the prior density and that have weights that take the new observation into

account. We can now estimate the new posterior density at time $(i+1)\delta$ from this information. One can surely estimate from these positions and weights the most likely state of the system given the observations.

One cannot use the same samples over and over and at the beginning of each step one has to resample the new density P_{i+1} or else the algorithm goes haywire. This should be obvious here: Some of the samples get very low weights after the new data have been taken into account; if the new densities are not resampled then after a few steps one is dealing with samples all of which have weights near zero. More generally, resampling is a key feature of many Monte-Carlo schemes.

This is the Bayesian filter. The samples are often called "particles" in the literature, and this filter is also known as a "particle filter."

4.7. References

1. G.I. Barenblatt, *Scaling*, Cambridge Univ. Press, Cambridge, (2004)
2. G.I. Barenblatt and A.J. Chorin, *A mathematical model for the scaling of turbulence* Proc. Nat. Acad. Sci. USA, 101, (2004), pp. 15023-15026.
3. A.J. Chorin, *Vorticity and Turbulence*, Springer, NYa, (1994)
4. A.J. Chorin and P. Krause, *Dimensional reduction for a Bayesian filter*, Proc. Nat. Acad. Sci. USA, 101, (2004), pp. 15013-15017.
5. C.K. Chui and G. Chen, *Kalman Filtering*, Springer, Berlin, (1987)
6. A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte-Carlo Methods in Practice*, Springer, NY, (2001)
7. B. Efron, *The Jacknife, the Bootstrap, and Other Resampling Plans*, CBMS-NSF Regional Conference Series, SIAM, (1982).
8. M. Ghil and P. Melanotte-Rizzoli, *Data assimilation in meteororology and oceanography*, Adv. in Geophysics, 53, (1991), pp. 141-256.
9. I. Gikhman and A. Skorokhod, *Introduction to the Theory of Random Processes*, Saunders, Philadelphia, (1965)
10. P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, (1992)
11. R. Miller, *Introduction to the Kalman Filter*, Sem. on Data Assimilation, European Center for Medium-Range Weather Forecasts, Reading, UK, (1997), pp. 47-60.

12. A. Yaglom, *An Introduction to the Theory of Stationary Stochastic Processes*, Dover, NY, (1962)